DCPAPERS

Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002: 35-43

Linking Collection Management Policy to Metadata for Preservation – a Guidance Model to Define Metadata Description Levels in Digital Archives

Maria Luisa Calanag, Koichi Tabata, Shigeo Sugimoto
University of Library and Information Science
Kasuga 1-2, Tsukuba City, Japan
{calanag, tabata, sugimoto} @ ulis.ac.jp

Abstract

In an environment of rapid technological change, collection managers face the challenge of ensuring that valuable resources remain accessible when there are changes to the technological context in which those resources are embedded. In this context of requiring "accessibility over time", digital preservation initiatives also demand for interoperability, or as what Hedstrom calls temporal interoperability. But first, libraries, especially in the academic world, need some general guidelines to assist in selectively choosing digital resources which are of great need to collect and preserve. This paper attempts to provide some structure for the concepts and ideas on a general collection management decision guide in the form of a requirements analysis framework that may assist in determining the metadata granularity required for digital resource management within an archive. The objective is for metadata and mechanisms to be shared among digital archives, but policies can be tailored to the requirements of the organization.
Keywords: *digital archives, preservation metadata.*

1. Introduction - Interoperability over time

We need scalable tools and standards for interoperability between archives.
Margaret Hedstrom

Paul Miller of UKOLN's Interoperability Focus, defines "interoperability" as follows: "to be interoperable, one should actively be engaged in the ongoing process of ensuring that the systems, procedures and culture of an organization are managed in such a way as to maximize opportunities for exchange and re-use of information, whether internally or externally".

1.1 Layers of interoperability

To achieve interoperability, the most practical way is to comply with standards. However, implementers often have to choose between standards and how to apply these high-level principles and standards to the "real world". From a "layer" model view, technical interoperability might be seen as the base on which other layers are built, where XML is seen as the standard facilitating technical interoperability. On the other hand, initiatives such as the Dublin Core Metadata Initiative (DCMI) and the Resource Description Framework (RDF) are seen as facilitating semantic interoperability (Johnston, 2001).

Hedstrom (2001) describes the concept of temporal interoperability as the ability of current systems or legacy systems to interoperate with future systems that may use new formats, data models, languages, communication protocols, and hardware. Temporal interoperability promises to make the digital archives of the future as interoperable as today's digital libraries.

Johnston (2001) further mentioned that there is also the aspect of inter-community interoperability that has to be considered, and that "collection description" could be a mechanism to attain this type of interoperability. Libraries have traditionally concentrated on the description of the individual items of their holdings in the form of catalog records. In a networked environment, there is a growing recognition of the value of complementing this item-level description with descriptions of higher-level aggregates of material. Collection descriptions can give an overview of groups of otherwise "uncataloged" items. Managers of archival and museum resources have traditionally made greater use of description at higher levels. As one example, the RSLP Collection Description project

https://doi.org/10.23106/dcmi.952106928

developed a collection description metadata schema which was based in part on the Dublin Core metadata element set, and an RDF implementation of that schema. The RSLP schema can be used in our proposed model. Sections 2.3 and 2.4 discuss more about collection-level descriptions.

1.2 Existing standards

Preservation metadata is comprised mainly of what most people would refer to as descriptive, administrative and structural metadata. There are a huge number of metadata initiatives, and it is difficult to figure out how these initiatives can work together. Dale (2002) explained how initiatives like the Open Archival Information System (OAIS), the Open Archives Initiative (OAI), the Online Information eXchange (ONIX), and the OpenURL could potentially fit and work together in the following ways. OAIS provides a general framework and an information model, with the U.K.'s Cedars project leading the way in developing an OAIS-based metadata specification. The OAI standards, on the other hand, defined ways in which descriptive metadata (Dublin Core) could be shared between organizations. ONIX is a standard for the representation and communication of product information from the book industry. OpenURL is a standardized format for transporting bibliographic-type metadata between information services and could be used as a basis for reference linking. It is possible that an OpenURL could link to an OAIS Dissemination Information Package (DIP). As for a 'wrapper' that would be able to link content and all this metadata together, there is now the XML-based Metadata Encoding and Transmission Standard (METS) initiative, now coordinated by the Research Libraries Group (RLG). METS is one option for encoding all of the information that would make up OAIS Information Packages. METS objects and OAIS Archival Information Packages (AIPs) would contain inside them all of the other types of metadata.

The METS schema builds upon the work of The Making of America II project (MOA2) and provides an XML document format for encoding metadata necessary for both management of digital library objects within a repository and exchange of such objects between repositories. A METS document consists of four main sections: Descriptive metadata, Administrative metadata, File groups, and Structural map. The European Commission co-funded Metadata Engine (METAe) Project , for instance, decided to adopt the METS schema as its standard output schema for several reasons. Firstly, METS emerged from the MOA 2 project, hence, it has a strong practical implementation aspect. Second, it has an open and flexible structure. Third, it is publicly available at the Library of Congress, and most of all, it is a well-described schema.

1.3 Scope and some challenges of web preservation

Since digital libraries are dynamic and widespread, with content, structure, location, delivery systems, and users changing frequently and instantaneously, they require new thinking and models for information management, access, use, and long-term archiving and preservation (Griffin, 2000).

Traditional libraries stress:

- Service
- Selection, organization, structure for access
- Centralization, standards
- Physical objects and standard genres

Contemporary technological capabilities (e.g. WWW) stress:

- Flexibility, openness
- Rapid evolution
- Decentralization (geographic, administrative)
- Digital objects, old and new genres

Digital preservation and digital archiving have been used interchangeably. Both terms mean taking steps to ensure the longevity of electronic documents. The 1996 Task Force Report on Archiving of Digital Information produced by the Commission on Preservation and Access (now the Council on Library and Information Resources) and the Research Libraries Group (RLG) considers long-term preservation as similar to archiving, and actually identifies digital archives, rather than digital libraries, as the unit of activity for the long-term preservation of digital materials. How does a library differ then from an archive? In the traditional sense of the word, these two institutions are usually distinct and separate entities with libraries focusing on the access function, and archives concerned with preservation. In the networked environment though, it would seem that archives are considered worthless without an access functionality or service.

The Internet Archive, for instance, started out simply as an "archive" according to the definition above. It attempted to collect all publicly accessible web pages, and these were "dumped" into a computer system with no organization or indexing. Even then, the fact is that without the vision of Brewster Kahle and his project's automated approach, these web materials would already have been lost. The nice thing is that an "Internet Library" service has been launched by the Internet Archive in 2001 through its Wayback Machine which now allows people to access archived versions of web sites, although it is still not a perfect system.

To be able to preserve web publications, it is necessary to know the construction of the web and some definitions used to describe it. The web is a way of viewing pieces of information located in different places on the Internet as if they were one large indexed document by using hypertext and multime-

https://doi.org/10.23106/dcmi.952106928

dia technique. This means that in a way it is impossible to preserve single publications completely because they have links pointing to other documents, which in turn link to others. Long-term preservation of the web seems to be hard to achieve, since a web page could not be preserved on paper or microfilm because the hypertext and multimedia techniques embedded will get lost and can never be retrieved again. Hence, the authors are also interested and concurrently looking into some ways on how to maintain this link functionality "over time" with the use of metadata.

1.4 A requirements analysis framework for formulating metadata guidelines for collection management & preservation

Collection management policies that deal with digital materials present one of the most critical challenges collection managers have to face. This will not be limited to technical issues only, but equally important are the organizational and management issues. Preservation decisions must be done at an early stage of the lifecycle of resources, since delays in taking preservation decisions can later result in preservation requirements that are more complex and labor intensive. Therefore, there is a strong need to establish guidelines that can assist collection managers in recording the appropriate level of metadata for collection management and preservation. The goal of this paper is to offer a requirements analysis framework which associates collection management policy with metadata to help collection managers define appropriate metadata based on their own requirements. The desired effect is for collection-level metadata and mechanisms to be shared among digital archives, but policies can be tailored to the requirements of the organization.

2. Collection management&preservation

"The next great revolution in libraries will be in collection development."
Stanley Chodorow

"Collection management policy" is a broader term than collection development, intended also to include storage, maintenance, de-selection and preservation. It is an important tool for defining what materials are of long-term interest to the collection. It needs to specify the acceptable level of functionality that has to be preserved if a digital object is to be retained. Such decisions will influence the level and method of access that will be necessary for the object as well as the level of preservation metadata required for long-term retention. For digital materials, value judgments made by the archivist/collection manager will determine what level of functionality needs to be retained. The Cedars Project has coined the term "significant properties" to describe those components of a digital object deemed necessary for its long-term preservation. Determining the significant properties of a digital object, i.e. the acceptable level of functionality, will dictate the amount of information or "metadata" that must be stored alongside the bytestream (the Data Object) to ensure that it remains "renderable" over time. How much specificity can be added to the metadata description, while maintaining broad applicability at the same time – is the authors' motivation in proposing the use of metadata description levels in this paper.

2.1 The responsibility for web preservation

Digital preservation is defined as the managed activities necessary for ensuring the long-term maintenance and continued accessibility of digital materials. It involves two main functions: the long-term maintenance of a bytestream and continued accessibility to its contents. Effective lifecycle management depends on a proactive approach and the cooperation of a number of stakeholders including content creators (See Figure 1 for the lifecycle diagram).

According to Allen (2000), "The management of digital collections is becoming a core Web-based service". He used the acronym SOAP to describe his essential elements in collection management, which are Selection, Organization, Access, and Persistence. It is also realistic to assume that we can not depend on data creators to preserve their own work because they lack both the power and the motivation to do so. Casey (1998) points out that the creator is rarely the "owner" of the server space where a site is located:

More often than not, Web site stability relies on who "owns" the server space where a site resides. Ownership means that the author of the Web site has control over use of the space as long as the content is within the policies of the administration of the server... Many folks in the academic world use the space allowed them on their university accounts to post Web pages. They cannot claim ownership of this space, just the right to borrow it for as long as they are associated with the institution or according to the Internet usage policy of the university. The irony is that many of these sites possess the content and quality that librarians want to preserve.

Preliminary results of a survey (Greenstein et al., 2001) issued by the Digital Libraries Federation (DLF) to its members discussed the library's relative role in creating, providing access to, and preserving digital assets within the university that contribute new forms of scholarly communication (e.g. e-journals, e-print repositories, digitized content, etc.). Many units within the university are taking responsibility for the production of digital content that contribute new forms of scholarly communications. The

https://doi.org/10.23106/dcmi.952106928

library is primarily responsible for the production of that content based on library holdings. Responsibility for other such content is widely spread across units with academic departments taking primarily responsibility for e-print repositories, e-journals, and distance learning materials. IT and academic computing departments have limited responsibility for production of digital information content of any kind.

The library though has a greater role in providing access to this content much more than creation of content. It is primarily responsible for providing access to digitized library content, to e-journal content, to e-books and to e-prints. Where preservation of such content is concerned, only the digitized library holdings appear at all to be secure. Most respondents to the DLF survey claim that the library takes responsibility for the preservation of these holdings, but other kinds of digital content such as e-journals and e-prints are apparently at risk.

2.2 Lifecycle management of digital materials

In traditional records management, the term 'information lifecycle' has long been used to describe the processes related to the creation and management of information. This concept is illustrated in Figure 1 (Brown, 2000). Preservation of digital materials needs to be an integral part of digital collection management and must therefore be incorporated into the overall management of an organization from acquisition through to preservation. It requires active management that begins at the creation of the material and depends on a proactive approach by digital repositories and the cooperation of stakeholders.

2.3 Collection descriptions

In the library domain, discussion has tended to focus on so-called "item" level metadata (i.e., descriptions of individual books, articles, and so on). The new environment brings about new requirements. The broker needs to have access to various types of metadata to support its operation. This is data about its environment and the resources in it. It should be clear that metadata is of central importance in distributed information environments.

Typically information objects exist in collections, where a collection comprises similar information objects. These collections might be databases, websites, document supply centers or libraries. They may be particular collections within a library, or the catalog for such collections. Such collections are also, of course, information objects, and collections may contain other collections. Collections will also have different terms and conditions associated with their use. Typically collections will be managed by organiza-

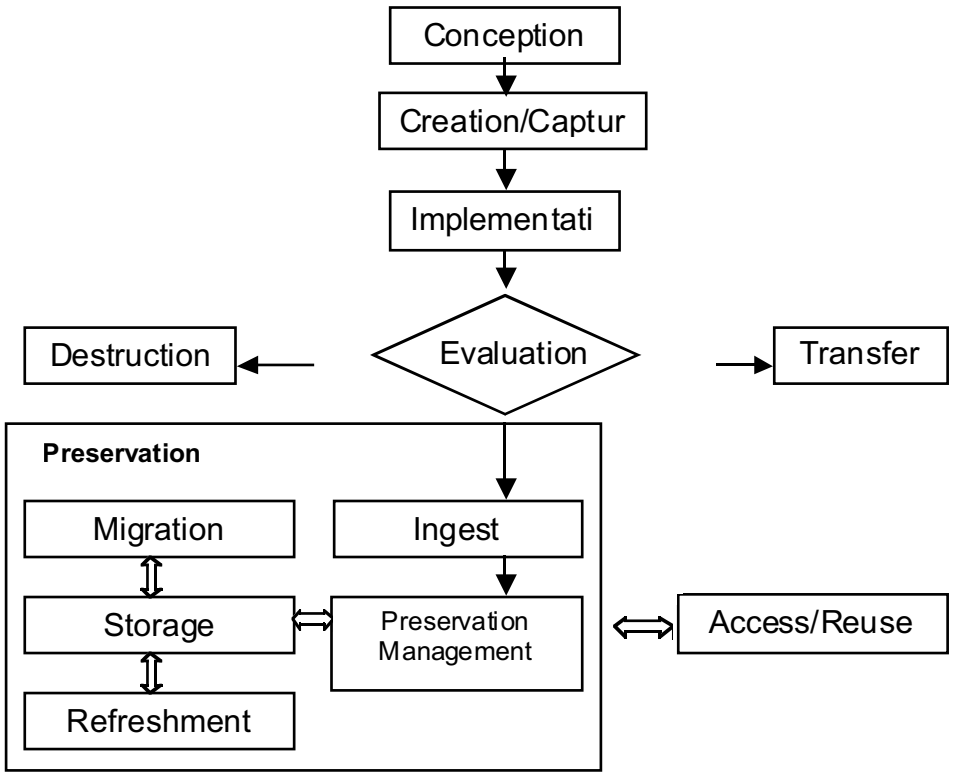


Figure 1. The Information Lifecycle (Used with permission from the Cedars Project. Cedars Guide to Digital Collection Management, 2002)

https://doi.org/10.23106/dcmi.952106928

tions. Information objects may be data or metadata.

Those who used Dublin Core (DC) expressed a strong need for item-level access, and somewhat less concern for grouping items into definable collections or sub-collections. There seemed also to be more uniformity of size and type among their materials. DC is cheaper to work in than MARC because of its limited element set. Those who used the Encoded Archival Description (EAD) standard wanted to organize items into collections and sub-collections, and saw the items just like chapters in a book or articles in a journal. A lack of uniformity of size and type within a collection also made EAD attractive. EAD is also cheaper to work in than MARC, because large numbers of items can be grouped (Seadle, 2001).

Archival description is the equivalent in archivology to cataloging in librarianship. There are important differences of principle and practice between these two fields. The definition from the General International Standard Archival Description (ISAD(G), 2001) makes use of two important concepts underlying archival management, which are as follows:

- The principle of representation: Because original archival materials cannot be organized for direct physical access by users, they have to be managed and retrieved by using representations. These have to contain the right data to allow for their effective use in the various management functions.
- The unit of description: The basic unit in archival management is taken to be the group ("fonds" in international usage, and also often called a "collection"). Most often, a group is a large body of materials that can be subdivided into subordinate entities. It would be normal, therefore, for an archive group to have a description representing the whole group, followed by a number of interlinked descriptions of its components. Generally, archival descriptions must contain information on the provenance, background, and context of the materials. It is, in principle, not possible to describe archival materials in terms of their contents and physical form alone. Provenance information includes a history of the administration or activity that caused the archives to be created and explains how they were used during the period when they were current records.

2.4 Linking policy to metadata

By merging traditional collection levels (Table 1) and collection level descriptions for digital resources which we call "Persistence levels" (Table 2) in the form of a matrix (Table 3), this can serve as a good starting point for developing a method of linking policy to metadata (Calanag et al., 2001). In addition, a set of values can be chosen for each combination according to the degree to which digital materials are persistent based on LeFurgy's (2002) definitions. Persistence is based on consistent and transparent rules for description and structure, standardized file formats, and so forth. In general terms, LeFurgy said that degrees of persistence can be represented in three categories (LeFurgy, 2002). In Table 3, these confidence ratings are what we considered as "Preservation requirement levels" in this paper.

- High (H): Fully persistent materials that enable high confidence for ongoing preservation and access.
- Medium (M): Partially persistent materials that enable medium confidence for ongoing preservation and access.
- Low (L): Marginally persistent materials that enable low confidence for ongoing preservation and access.

Given that persistence is closely tied to the clarity and consistency with standards by digital resources, it follows that materials that are highly structured tend to be inherently easier to preserve and access over time. Conversely, less structured materials tend to be harder to manage. In addition, persistence can also be tied to resource availability in terms of the digital object's persistent identifier.

The authors propose that these three Preservation requirement levels (High/Medium/Low) may determine the granularity of the preservation metadata that will be required to ensure that the digital materials will be preserved and accessed over time. In other words, a choice among High/Medium/Low can be associated with item-level, class-level, or collection-level preservation metadata, respectively (see Table 4). As shown in a Sample Policy Table (Table 3), a general rule of thumb is that we go from High to Low as the persistence levels gain lower confidence and stability. Collection manager-defined default ratings or a blank space(s) denoting Not Applicable can be assigned according to the institution's policy.

3. Digital archives in academia

"Universities are becoming publishers and they need to take responsibility for their own output."
Cedars Final Workshop summary

There is some need for institutional responsibility from universities, especially with regard to local scholarly material, learning objects and institutional records. Cedars, for example, had focused on the incoming digital acquisitions of research libraries and the intellectual content created by institutions, both digitized and "born digital". Preservation was about the continued accessibility of the content of digital resources, and was focused on the content rather than any particular medium. One major difference between traditional collection management strategies and that needed for digital information is that consideration of preservation requirements needed to hap-

https://doi.org/10.23106/dcmi.952106928

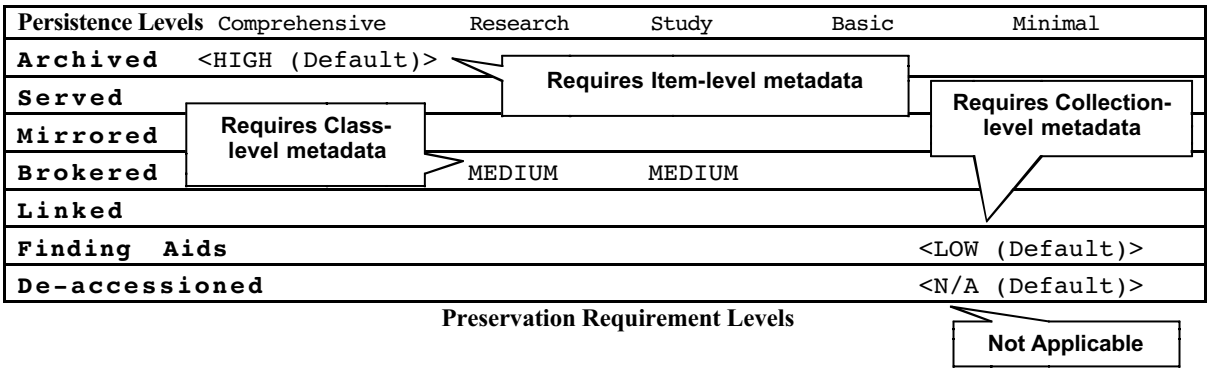
Table 1. Collection levels

Levels	Description
Comprehensive	A collection to include all significant works of recorded knowledge in all applicable languages for a defined and limited field.
Research	A collection which includes the major dissertations and independent research, including materials containing research reporting new findings, scientific experimental results, and other information useful to research.
Study	A collection which is adequate to support undergraduate and most graduate course work, and to maintain knowledge of a subject required for limited or general purposes.
Basic	A highly selective collection which serves to introduce and define the information available elsewhere.
Minimal	A collection in which few selections are made beyond very specific works.

Table 2. Persistence levels

Levels	Description
Archived	Material is hosted in the library, and it intends to keep intellectual content of material available permanently.
Served	Material is hosted in the library, but no commitment to keeping it available.
Mirrored	Copy of material residing elsewhere is hosted in the library, and it makes no commitment to archiving. Another institution has primary responsibility for content and maintenance.
Brokered	Material is physically hosted elsewhere and maintained by another institution, but the library has negotiated access to it; includes metadata and links in the catalog, and library users can locate and cross-search it.
Linked	Material is hosted elsewhere, and the library points to it at that location; no control over the material.
Finding Aids	Electronic finding aids and metadata held by the library to facilitate discovery and searching; this metadata is associated with the library's digital collections or elsewhere, but may be stored, managed and maintained separately from them.
De-accessioned	Accessioned resources that have not been retained after review.

Table 3. Putting it all together:
A Requirements analysis matrix linking policy and metadata – A Sample Policy Table



In using this matrix, a general rule of thumb is that we go from High to Low as the persistence levels gain lower confidence and stability. Collection manager - defined default ratings or Not Applicable <N/A>ratings can be assigned according to the institution's policy.

pen much earlier in a resource's life cycle. Decisions taken at each stage in the lifecycle would influence options at other stages. It follows, therefore, that creators play a significant role in digital preservation.

The most likely collection model would be distributed, but there would be a need for transparency as

https://doi.org/10.23106/dcmi.952106928

to which organizations are preserving what materials and clarification of roles and responsibilities. We would have to adapt to high volumes of information which would stress the importance of distributed solutions and the automation of ingest and metadata capture processes. There would also be a need to find and manage the information, based on metadata and persistent identification. Another major challenge in the academic sector would be e-prints and e-theses. The scale of the challenges faced would mean future archiving would be distributed.

3.1 Persistent archive architecture

Archivists rely on persistent archives to support all aspects of data collection management. Persistent archives provide a mechanism needed to support distributed data access across heterogeneous data resources (Moore, 2002; Ludascher, et.al., 2001). Using concepts and terminology from the Open Archival Information System (OAIS) reference model, Figure 2 shows a digital archive architecture that can be built around XML-based standards and technologies.

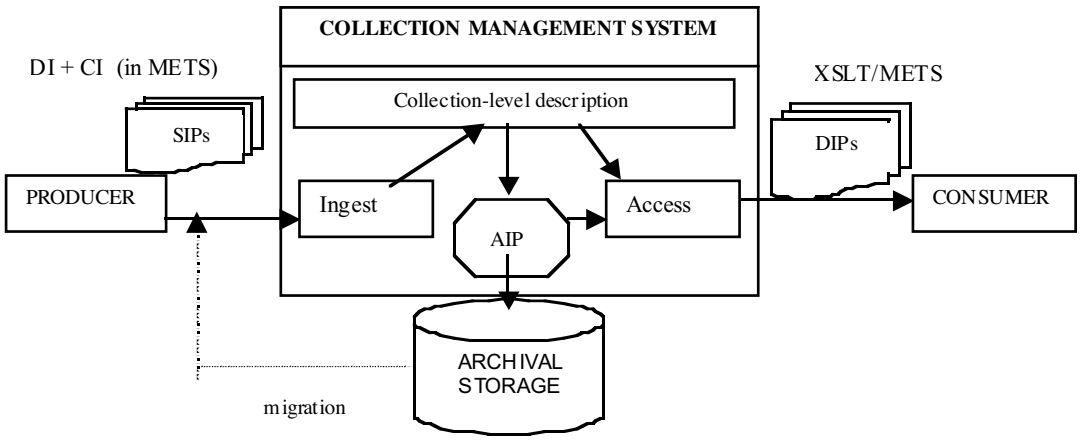
First, the producer and the archive need to agree on the submission policies (e.g., acceptable submission formats, specifications on what are to be preserved, access functions, and other legal requirements), and the preservation policies. General preservation decisions can be made based on the matrix presented in Table 3 which will serve as a requirements analysis framework. Then, the producer can ingest these SIPs (METS-encoded Submission Information Packages = Descriptive Information + Content Information) into the Collection Management System where they are assigned the appropriate metadata at the granularity level based on the requirements analysis framework. The "highly persistent" (H) the resource is considered to be, the more detailed preservation metadata should be to allow the resource to be emulated, for example, on future platforms. Once these additional information are added to Descriptive metadata, SIPs are transformed into AIPs (Archival Information Packages) which are put into archival storage. Migration of AIPs in the archive is simply a regular refreshing process (for now) to prevent obsolescence of AIPs. The bitstream of the content will remain unchanged. Only new provenance metadata will be added every time medium migration is done.

The levels of metadata granularity are described in Table 4 which shows their equivalence to the preservation requirement levels.

3.2 Preservation metadata at the three granularity levels

Figure 3 presents a simple collection description model to provide a view of the framework into which the metadata granularity level fits. Most of the preservation metadata elements enumerated in Appendix 1 have been recommended by the OCLC/RLG Working Group on Preservation Metadata (2002) in their latest report. Grouping the metadata elements according to the three granularity levels, is one possible categorization proposed by the authors.

This is how the proposed "collection management decision guide" (Table 3) can be applied. Default ratings can be set for certain combinations. However, let us take a specific example, a HIGH rating has been assigned to the combination SERVED + STUDY by the collection manager. This means that Item-level description or metadata should be provided for each



Legend:
DI - Descriptive Information
CI – Content Information
SIPs – Submission Information Package(s)
AIP – Archival Information Package
DIPs – Dissemination Information Package(s)

Figure 2. Digital archive architecture

https://doi.org/10.23106/dcmi.952106928

Table 4. Mapping between Preservation requirement levels and metadata granularity

	Metadata granularity	Description
High	item-level metadata	Individual digital objects are packaged into the Content Information (CI).
Medium	class-level metadata	Structural information is handled; this metadata describes types of object attributes, and aggregation information (Context Information)
Low	collection-level metadata	Can be added to the Descriptive Information (DI) and in this paper, this also refers to the RSLP collection description schema

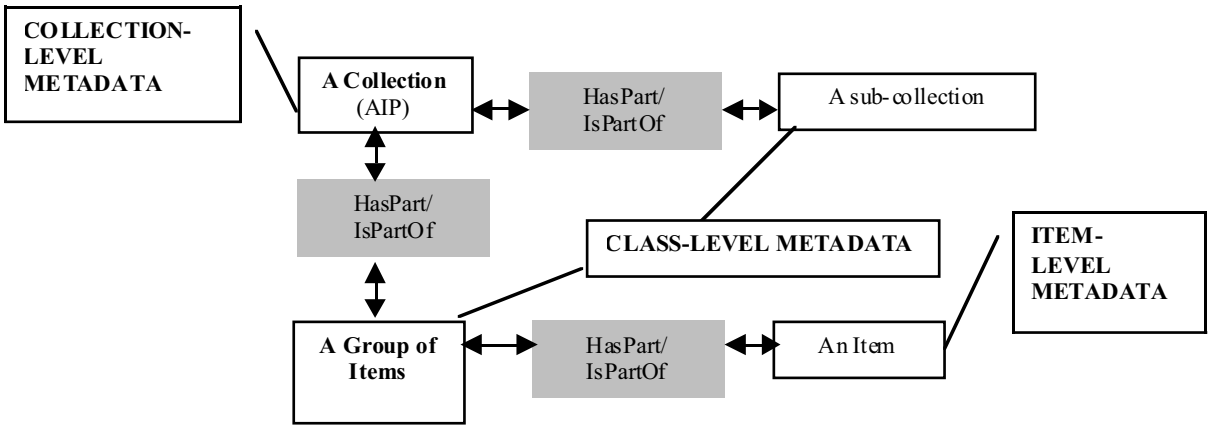


Figure 3. Collection description model

item in the aggregation or set. To designate a HIGH rating entails a big responsibility and commitment on part of the institution since very detailed metadata has to be provided or generated (see Appendix 1). Decisions have to be given much thought by the collection manager, and one main criterion that can guide decision-making is the persistence of materials. On the other hand, if it has been decided that LINKED + BASIC =LOW, then it means that Collection-level description or metadata should be used. These collection-level descriptions or metadata can then be shared among digital archives for cross-searching, access and re-use.

4. Conclusions

The authors have laid down a collection management guide in the form of a requirements analysis matrix for general applicability in the academic environment, where preservation policy decisions can be made according to local requirements. It also prescribed a digital archive architecture that can be used in distributed environments which can serve as a mechanism for institutions to coordinate their digital preservation activities while at the same time, retaining the flexibility to meet their local needs.

In selecting materials for preservation, evaluation decisions might reflect technical issues including the quality of the data object and its existing metadata, and the technical environment, both hardware and software, needed to access and use the data object. According to the persistence of resources as determined by collection managers and/or information producers, this paper prescribed a way to ensure that documentation will be preserved so that environments can be reconstructed for future "processability" or accessibility.

For organizations taking responsibility for the long-term preservation of digital materials, a written and up to date collection management policy is critical. It provides an important tool for the collection managerby inviting consideration of all the relevant issues early in the lifecycle of digital materials within their scope.

Two vital criteria for preservation are to ensure that the preserved digital object can be found, and that the preserved digital object can be understood. For these criteria to be met, it is vital that each preserved digital object has a unique and persistent identifier. For their future work, the authors are currently conceptualizing a mechanism for encoding preservation metadata in a URL that offers context-sensitive links that should lead to the appropriate versions of the resource that the user needs.

https://doi.org/10.23106/dcmi.952106928

References

Allen, Robert B., 2000. Digital libraries - a cornerstone of web-based services. *In*: International Conference of Asian Digital Library 2000, Seoul.

Brown, Adrian, 2000. Digital Archiving Strategy. English Heritage. http://www.english-heritage.org.uk/default.asp?wci=WebItem&WCE=544

Calanag, Maria Luisa, Sugimoto, Shigeo, and Tabata, Koichi. A metadata approach to digital preservation. *In*: *Proceedings of the International Conference on Dublin Core and Metadata Applications 2001. Tokyo: National Institute of Informatics, October 2001*.

Casey, Carol, 1998. The Cyberarchive: a look at the storage and preservation of web sites. *College and Research Libraries* , 59 (4), 304-310.

Chodorow , Stanley, 2001. Scholarship, Information, and Libraries in the Electronic Age. *In*: Marcum, D.B., ed. *Development of digital libraries: an American perspective*. Westport, Conn.: Greenwood Press.

Dale. Robin, 2002. Cedars final workshop. Manchester: UMIST.

Day, Michael and Jones, Maggie, 2002. Cedars (CURL Exemplars in Digital Archives) final workshop: executive summary. Manchester: UMIST.

Greenstein, D., Thorin, S. and McKinney, D., 2001. Draft report of a meeting held on 10 April in Washington Dc to discuss preliminary results of a survey issued by the DLF to its members.

Griffin, Stephen M., 2000. The Digital Libraries Initiative: a USA federal program of research and applications. *In: Digital Libraries*. No. 18.

Harrod's Librarians Glossary, 1999. Compiled by Ray Prytherch. Gower.

Hedstrom, Margaret, 2001. Exploring the concept of temporal interoperability as a framework for digital preservation. University of Michigan.

ISAD (G): General International Standard Archival Description. Ottawa: International Council on Archives, 2000.

Johnston, Pete, 2001. Interoperability: supporting effective access to information resources. *Library & Information Briefings*. London: South Bank University.

LeFurgy, William G, 2002. Levels of service for digital repositories. *D-Lib Magazine*.

Ludascher, B., Marciano, R., and Moore, R., 2001. Preservation of digital data with self-validating, self-instantiating knowledge-based archives. *SIGMOD Record*, vol. 30, no. 3, pp. 54-63.

The Making of America II Testbed Project. http://sunsite.berkeley.edu/MOA2/moaproposal.html

Metadata Encoding and Transmission Standard (METS): an overview and tutorial. http://www.loc.gov/standards/mets/METSOverview.html

Moore, Reagan W., 2002. Persistent Archive Concept Paper Draft 2.0. San Diego Supercomputer Center: Persistent Archive Research Group Global Grid Forum.

The OCLC/RLG Working Group on Preservation Metadata, 2002. Preservation metadata and the OAIS information model: a metadata framework to support the preservation of digital objects. Ohio: OCLC Online Computer Library, Inc.

The Prism project: vision and focus. Prism working paper 2000-02-07 http://prism.cornell.edu/Publications/WorkingPapers/Visions.htm

Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems. May 1999. http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

Seadle, Michael, 2001.Sound practice: a report of the best practices for digital sound meeting, 16 January 2001 at the Library of Congress. *RLG DigiNews*, April 15, 2001, vol. 5 no.2 http://www.rlg.org/preserv/diginews/diginews5-2.html

Task Force on Archiving of Digital Information. Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. 1996.

https://doi.org/10.23106/dcmi.952106928